Examining control of a multi-scale system: Using Deep Reinforcement Learning to control an agent-based model of sepsis

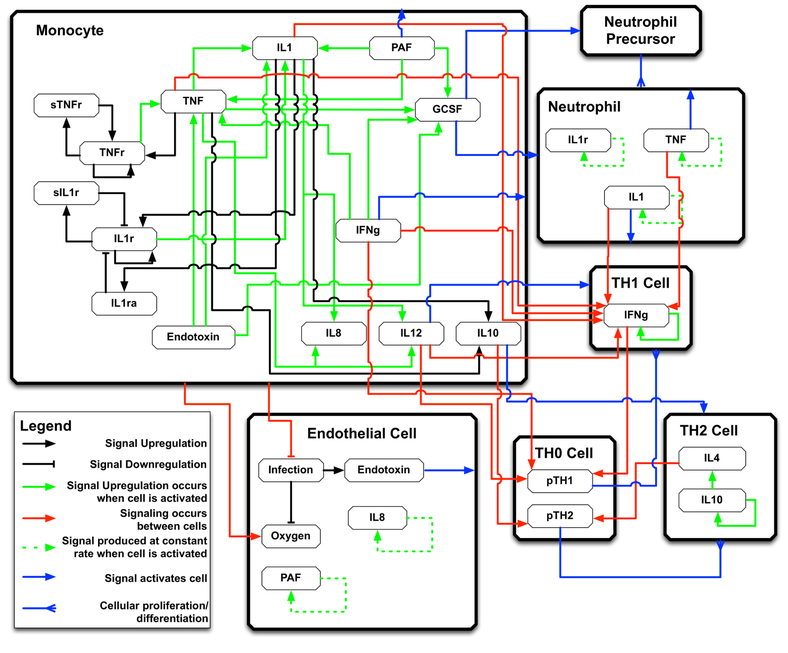
Brenden Petersen[1], Jiachen Yang[1], Will Grathwohl[1], Claudio Santiago[1], Chase Cockrell[2], Gary An[2] and Dan Faissol[1]
[1] Lawrence Livermore National Laboratory; [2] Department of Surgery, University of Chicago

Introduction

Effective treatment of complex diseases, such as sepsis, cancer, diabetes, and chronic inflammation require correspondingly complex control strategies involving multiple targets/levers, the configurations of which must vary both from patient to patient as well as during the time course of a single patient. We propose that the use of deep reinforcement learning (DRL), on simulations of pathophysiological processes can guide the development of multi-modal and adaptive therapeutic regimens. We present an initial example of DRL to identify an adaptive control policy for controlling the Innate Immune Response ABM (IIRABM) as a proxy model of sepsis. This work demonstrates a path forward in terms of controlling complex disease while fully leveraging advanced machine learning/AI methods.

Deep Reinforcement Learning: Means of Adaptive Control

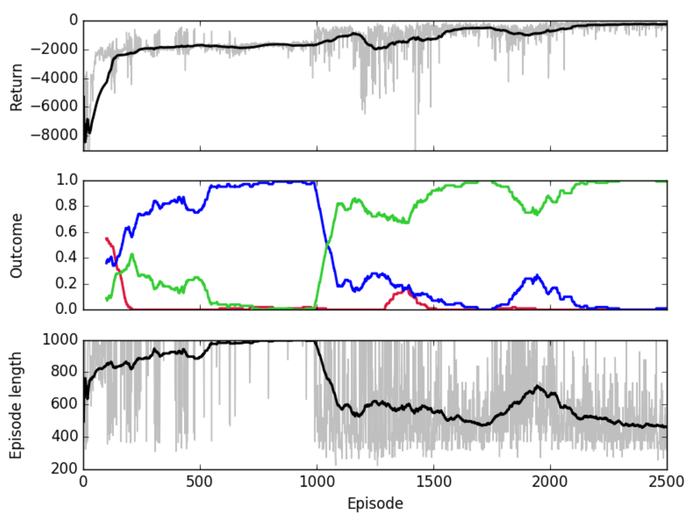
- Machine learning/artificial intelligence approach to "playing games" => e.g. Pong, Lunar Lander, **AlphaGo**
- Train a ***policy*** that consists of an optimal response/ control ***action*** based on the ***observed*** state of a system
- Iterative learning process => system calculates optimality of an action based a defined ***learning objective***
- ***Reinforcement Learning*** => Uses a matrix/lookup table to calculate effect of actions => Limited capability/ applicability to complex systems
- ***"Deep" Reinforcement Learning*** => Trains a neural network to identify optimal policy = **DRL Agent**
- What makes it "adaptive:" specific sequence of controls varies based on individual system trajectory; policy is a sum of actions given a particular observed state
- ***Requires a simulation of the system to train on => Innate Immune Response ABM as a proxy for sepsis***



Training and Learning

Training Conditions of the **DRL Agent**:

- Fully Observable System: 21 variable/grid point x 101x101 grid every time step (6 mins)
- Discrete Action Space for +/- 14 cytokines per Time Step (6 minutes sim time)
- Reward Function:
  - Terminal = Life vs Death
  - Intermediate = Redution of system damage
- 1 Episode = 1 simulation run
- Trained on single parameterization (stochastic replicates) with 46% mortality



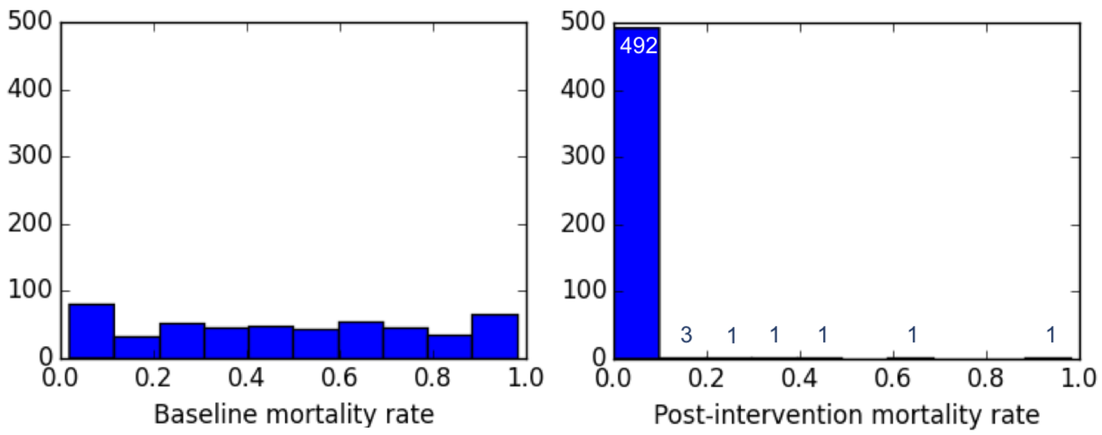
- **Top Panel: Return per episode (grey = single episode, black moving average/100 episodes)**
- **Middle Panel: Moving average over 100 episodes of rates of patient outcome (death = red, timeout = blue, health = green)**
- **Bottom Panel: Episode length in steps (grey = single episode, black moving average/100 episodes)**

Result: A Robust Policy

Test of Generalizability: Since **DRL Agent** trained only on 1 parameterization, is learned policy translatable to other initial conditions

- Tested over 500 different initial conditions, batched 100/10% mortality intervals, 500 stochastic replicates/initial condition

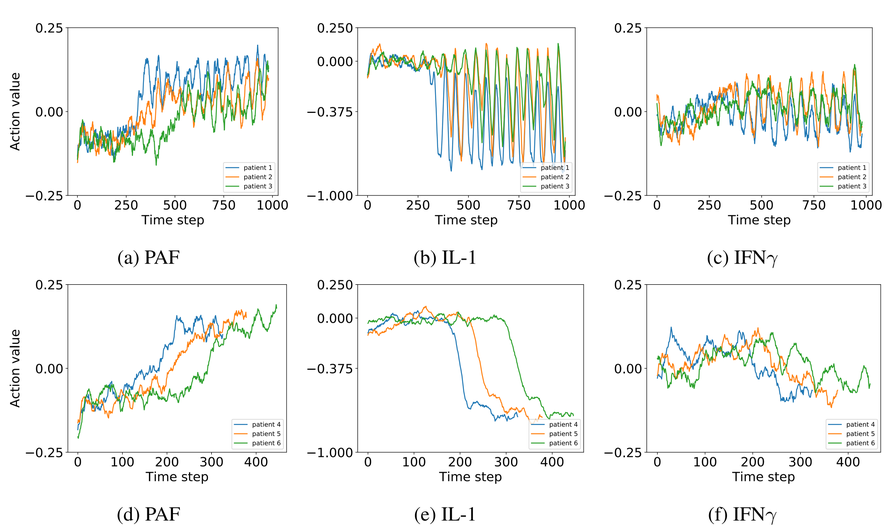
Results: Of 500 conditions, 92% had 0% mortality, 7.8% had reduced mortality (ave = 87% reduction), 1 condition unchanged, none worsened



Result: Adaptive Control

Principle of Adaptive Control = Different trajectories require different sequence of control actions

- Test 1 (Upper Row): 3 Varied levesl of recurrent injury (nosocomial risk)
- Test 2: (Bottom Row): 2 Varied levels of initial infection (injury severity)
- Panels show implemented actions in terms of degree of cytokine manipulation
- Action Space is biologically plausible:
- Test 1: Oscillating Containment of Nosomical insults
- Test 2: Prolongation of control w larger injuries, early suppression and late augmentation of PAF



Future Directions

Current Finding: DRL can be used to discover a Robust Control Policy for a complex, stochastic system

Next Steps:

- Modify Observation and Action space to define minimally sufficient intervals, targets and effects
- Optimization Problem/Active Learning Search to identify minimal criteria
- Current limitation of DRL = Interpretability => Extract biological insight from arrived at control policies, e.g. is there an "optimal trajectory" through multi-dimensional disease space
- Apply to other dynamic therapeutic targets (e.g. multi-modal chemo vs evolving resistance, wound healing in face of evolving wound microbiota)